2026 START Program
CFP Brief

THEME: **02.** **Agentic, Artificial Intelligence**

SUB-THEME: **2.4.** **Training AI-ISP Systems for Dynamic, Ambiguous, and Real-World Capture Conditions**

Context/ Overview

Modern smartphone imaging pipelines increasingly rely on deep learning-based components embedded throughout the image signal processor (ISP). Moreover, most modern ISPs now utilize multi-frame strategies where a burst of consecutive sensor frames are captured as input and fused to produce a single output. A major bottleneck in AI ISP development is the need for paired training data. In a multi-frame scenario, the burst input is paired with a single "ground truth" output. A significant challenge is capturing such paired datasets. Most existing datasets are limited to indoor and/or stationary environment where the burst frames exhibit little motion. In addition, existing datasets rarely address dynamic scenes (e.g., the camera and/or scene move during the burst capture) or challenging image conditions (e.g., low-light or high-dynamic range environments).

Problem Statement

Paired data generation for AI-based ISP training becomes fundamentally ill-posed in dynamic and challenging capture conditions. Local and non-rigid motion that happens during burst capture results in scene occlusion/disocclusion among the frames. In addition, exposure variation for HDR and rolling shutter effects can break the assumptions underlying traditional pairing, in particular, that there exists a single "ground truth" output. Finally, traditional image-based loss-functions typically rely on pixel-wise alignment in training data (input to ground truth) which often is not valid in a multi-frame scenario in challenging imaging conditions.

Objectives & Scope

This CFP seeks advanced model and training framework development for AI-based ISP systems operating on multi-frame RAW burst inputs under dynamic and challenging imaging conditions. The scope includes innovations in paired data formulations, where degraded or intermediate sensor measurements (e.g., RAW bursts) are associated with reference signals or constraints, as well as supervision methodologies, including loss functions, consistency objectives, or other training signals used when pixel-aligned ground truth is ambiguous or unavailable.

Preferred evaluation settings involve AI-ISP components trained and evaluated on RAW burst sequences, with explicit consideration of camera motion (e.g., hand motion), scene motion, exposure variation, occlusion, and multi-frame interactions.

The program targets high-impact, non-trivial problems with a 2–3 year research horizon and the potential for long-term impact on real-world smartphone imaging pipelines. Proposals should present a clear multi-year research vision, including the intended technical impact and how the work advances the state of the art. Within this vision, proposals must define a well-scoped first-year project plan, specifying concrete technical objectives, methodologies, and evaluation criteria.

Successful proposals will demonstrate relevance to smartphone-class imaging systems, address challenging failure modes such as motion, occlusion, low light, and wide dynamic range, and outline a credible path toward reproducible and scalable data generation or supervision.

Specific Topics & focus areas*

Addressing the proposed problem may include research in the following areas:

1. Multi-frame RAW burst pairing under motion and low light
2. HDR and exposure fusion with scene dynamics
3. Loss and supervision design for ambiguous ground truth
4. Occlusion-aware capture and training
5. Realistic 3D simulation and environment-based pairing
6. Automated and Instrumented Capture Pipelines

※ The topics are not limited to the above examples and the participants are encouraged to propose other original ideas.

END OF DOCUMENT